Exercises

1	2	3	4	5	6	7	8	9

Surname, First name

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Modelling, Uncertainty and Data for Engineers (CEGM1000)
Exam 22/23 Q1

Do not open the exam until given permission by the instructor!
(you may write your name and student ID)

The exam is 180 minutes. The table an overview. On the following pages, some questions have a specific box for you to answer: anything written outside the boxes will not be graded. Note that we have provided a lot of space for answers. The answer space size is not an indicator of how long we expect your answers to be! (shorter is generally better). Points indicate the relative amount of time expected to be spent for each question. Scratch paper is available to use during the exam, but will not be collected or graded. You may use pen or pencil, a scientific calculator and the attached formula sheet.
Don't forget to write your student ID and fill in the bubbles on the top right of this page. Good luck!

No.	Question	Sub-Q	Type	Points
1	Coding	a, b	SA	6 + 4
		c-g	MC (e is SA)	2 each
2	Probability		SA	6
3	Probability	a-e	Calc	2+2+3+3+4
4	Probability	a, b	MC	1 + 1
		c	Calc	3
5	Mathematical Modeling	a-c	MC	2+2+2
6	Numerical Methods	a, b	Calc	6 + 6
		c	SA	4
7	Observation Theory	a, b	Calc	6 + 6
8	Observation Theory	a	Calc	3
		b	SA	5
9	Stochastic Processes	a, b	Calc	3 + 3
		c, d	SA	5 + 5
		e	MC	2
		Total:		**105**

Part 1: Coding [20p]

6p **1a** Briefly describe the three main types of errors you encountered while working with Python:

- syntax errors
- exceptions
- logical errors

Add one example for each of them in your description. [200 words maximum]

4p **1b** The code below computes an average test score from a series of individual tests carried out on a generic item of a certain production line(e.g., car components or appliances). Each test yields an integer number from 0 to 10. If the average score is >= 8, and if no individual score is below 5, the item is accepted for further processing; otherwise it is discarded.

The code below implements this computation, but has three errors.
<u>Identify the 3 errors (use the line numbers as reference) and explain briefly how to fix them.</u>

In [ ]:
```
1  def compute_final_score(individual_scores):
2      '''
3          Compute the average score of an item and check for acceptance (avg. score >=8).
4          It exits as soon as an individual score is too low (<5), with a return statement.
5  
6          Parameters:
7              individual_scores: A list of individual scores (from 0 to 10), one for each test
8  
9          Returns:
10             a string reporting whether the item was accepted or discarded (and its average score)
11     '''
12     n_scores = len(individual_scores)
13     total_score = 0
14 
15     for score in individual_scores
16         if score < 5:
17             return 'The item is discarded, individual score too low.'
18             total_score += score
19 
20     avg_score = total_score / n_scores
21 
22     if avg_score>=8:
23         return f'The item is discarded. Avg. Score = {avg_score:.2f}'
24     else:
25         return f'The item is accepted. Avg. Score = {avg_score:.2f}'
```

2p **1c** Select which of the following 3 statements are TRUE concerning **assertions** and **exceptions**? (you can select more than one statement)

Remember, the generic assertion statement is defined as follows:

```
assert condition, message
```

☐ We use assertions to catch generic Exceptions that can be raised at runtime

☐ The message in an assertion statement is optional

☐ An assertion statement is equivalent to:

```
if not condition:
    raise AssertionError(message)
```

2p **1d** What will happen when running this piece of code?

```
In [ ]:
def compute_speed(distance,time):
    speed = distance / time
    return speed

try:
    speed = compute_speed(10,0)
except Exception:
    print("You cannot divide by zero!")

print("End of code!")
```

(select only one answer)

- (a) The code will not run due to a Syntax Error
- (b) The code will run, but it will stop at runtime because we did not catch the right type of Exception, which is ZeroDivisionError
- (c) The code will run, but it will stop at runtime because we did not catch the right type of Exception, which is ValueError
- (d) The code will run, and it will print "End of code!"
- (e) The code will run, and it will print "You cannot divide by zero!" followed by "End of code!"

Consider the code below, defining a generic Rocket class:

In [ ]:

```
1  import numpy as np
2  
3  class Rocket():
4      '''
5          This class simulates a rocket ship for a game or a physics simulation.
6      '''
7      def move_up(self):
8          # Increment the y-position of the rocket.
9          self.y += 1
10 
11     def __init__(self):
12         # Each rocket has an (x,y) position.
13         self.x = np.random.rand()
14         self.y = np.random.rand()
15 
16 my_rockets = []
17 for x in range(0,5):
18     new_rocket = Rocket()
19     my_rockets.append(new_rocket)
20 
21 for rocket in my_rockets:
22     print(rocket.y)
```

2p **1e** Identify the following (you may use the line numbers as a reference):

1. The constructor of the class
2. A method of the class
3. An object of the class
4. An attribute of the class

2p **1f** What will happen when running the code in lines 21-22?
(select only one answer)

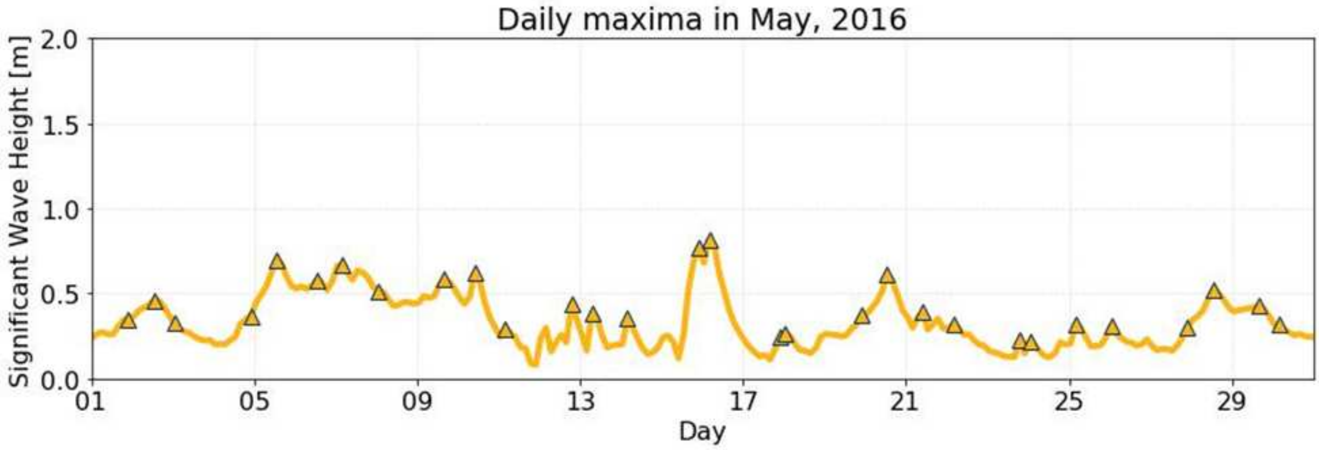
- (a) The code will not run, unless we replace rocket.y with new_rocket.y
- (b) The code will run and it will print the same random numbers multiple times
- (c) The code will run and it will print different random numbers
- (d) The code will run, but it will raise a NameError exception

2p **1g** Imagine the Rocket class is now contained in a module named space.py. Which of the following are correct import statements?
(you may select more than one statement)

- [ ] import space
- [ ] from space import Rocket as RocketFromSpace
- [ ] import Rocket from space
- [ ] from space import * as RocketFromSpace

Part 2: Probability [6p]

Consider the plot below, which is one month of hourly wave height measurements. You can assume this month is representative of conditions for a 5 year time series of wave heights.



6p **2** In the space below, describe the key steps that would be required to create a continuous parametric probability distribution that represents annual maximum wave height. State your assumptions, then list your analysis steps (be brief!), followed by a statement about whether or not there may be any limitations to this approach. An outline of what your answer should look like is shown here:

State assumptions (2 sentences max for this item)

1. *How to select maxima*
2. *Choose a distribution*
3. *How to check the distribution fit*

State limitations (2 sentences max for this item)

Part 3: Probability [14p]

X and *Y* are two (unit-less) quantities that have been measured in the lab in order to investigate certain properties. Rather than using a 'standard' parametric distribution, theoretical cumulative distribution functions for *X* and *Y* have been fitted satisfactorily to data obtained after many years of measurements. These are given by $F_X(x)$ and $F_Y(y)$ below:

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \frac{x+1}{2}, & -1 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 1 - e^{-y}, & y \ge 0 \end{cases}$$

2p **3a** What is $P(X \le -0.99)$?

A certain group of engineers want to design for values of *Y* with exceedance probability of 0.001. That is $P(Y > y) = 0.001$

2p **3b** What is the design value of *Y*?

Assume *X* and *Y* are independent.

3p **3c** What is $P(X \leq -0.1, Y \leq 1)$ (approximate using at least 4 decimal places)?

3p **3d** What is $P(X > 0.8|Y > 10)$?

As it turns out, the joint cumulative distribution function of *X* and *Y* (denoted as F_{XY}) has also been approximated from measurements with sufficient accuracy and is given:

$$F_{XY}(x,y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1} & -1 \le x \le 1, 0 \le y \le \infty \\ 1 - e^{-y} & 1 \le x \le \infty, 0 \le y \le \infty \\ 0 & \text{elsewhere} \end{cases}$$

4p **3e** What is $P(X > 0.8, Y > 4.6)$ in the case the joint distribution is as above (approximate using at least 4 decimal places)?

Part 4: Probability [5p]

A csv-file is used to create a pandas dataframe, df. The command df.describe() returns the following:

	RandomVariable
count	8760.0
mean	212.8
std	144.7
min	0.0
25%	130.0
50%	210.0
75%	280.0
max	990.0

1p **4a** Based on the data summary above, which of the following distributions would be a good first choice to represent the data?

- (a) Normal / Gaussian
- (b) Exponential
- (c) Uniform
- (d) Gumbel

1p **4b** Which of the following is a suitable justification for your answer to the previous question?

- (a) The median is bigger than the mean
- (b) The data seems to have a right skew
- (c) The data seems to have a left skew
- (d) I remember this from the sample exam
- (e) None of the above

3p **4c** You decide to use the Gumbel distribution to fit the data. Compute the distribution parameters using the data summary above

Part 5: Mathematical modelling [6p]

2p **5a** Assume you are given an assignment to model a specific system of interest (could be any system). To model it, you need to make some assumptions. Among the many criteria and constraints at stake, what is the main criteria that should drive your decision process to make such assumptions? (choose one answer)

- a Available time to develop the model
- b Available budget given, that covers the hours and any other expense of yours to actually develop the model
- c The purpose of the model, that should fit within a trade-off between complexity, affordability and accuracy
- d Data available to develop the model

2p **5b** When would you need to work on an inverse problem? (choose one answer)

- a Whenever you don't have enough data available to calibrate the model
- b Whenever you want to improve your model
- c Whenever you want to identify from measured data some unknown values of specific properties/parameters of your model
- d Whenever the inverse problem is well-posed and a unique solution is available.

2p **5c** As engineer and scientist, what should be our final goal after we develop a model? (choose one answer)

- a To validate the model
- b To verify and calibrate the model
- c To perform a sensitivity analysis on the model
- d To verify and perform a sensitivity analysis on the model

Part 6: Numerical Methods [14p]

6p **6a** Using Taylor expansion, derive the forward Euler approximation for the first derivative. Show the truncation error introduced by the approximation.

6p **6b** Derive the discrete form of the following ODE using the forward Euler approximation and calculate first 5 timesteps of the solution using dt = 0.2:

$y' = y + t\cos(t)$

$y(0) = 1$

4p **6c** A colleague proposes using the backward Euler method. List one advantage and one disadvantage of this approach compared to forward Euler.

Part 7: Sensing and Observation Theory [12p]

We can use two instruments to measure the water level at a given location and time. In the following, you may assume that all measurements are independent and the water level does not change between subsequent measurements. Instrument A has precision of 3 mm, instrument B has a precision of 8 mm.

6p **7a** It needs to be decided whether to take one measurement with the most precise instrument (option 1), **or** we take one measurement with each instrument and estimate the water level from both measurements (option 2). By how much will the precision of the estimated water level improve or deteriorate if option 2 is used instead of option 1?

6p **7b** Instrument A is expensive and complex to use, whereas instrument B is cheap and simple to use. One of the instruments must be selected for future use. Therefore it is assessed how many measurements are needed with each instrument to obtain an acceptable precision. What will be the 99% confidence interval of the estimated water level if we use 4 repeated measurements to estimate the water level with instrument A? And how many measurements do we need to take with instrument B to obtain at least the same (or tighter) confidence interval?

Part 8: Sensing and Observation Theory [8p]

Once per month the height of a fixed benchmark site on a volcano is determined using GNSS. The volcanologist uses as a null hypothesis that the height is constant. After 6 months (i.e., using 6 observations), she wants to test her hypothesis with the overall model test and a probability of false alarm of 0.025.

3p **8a** What would be the threshold value she needs to use?

5p **8b** Assume the null hypothesis is rejected. The volcanologist now wants to test whether a sudden height change occurred at a certain time t_c after the first observation at t_0 (i.e., height is constant before t_c, then changes at t_c and after that remains constant again). Specify the hypotheses and describe a testing procedure which allows to identify whether such deformation occurred and at which time it did. *No need to do any calculations.*

Part 9: Simulation and Stochastic Processes [18p]

We are interested in modelling the local weather and decide to build a discrete-state, discrete-time Markov chain model for this purpose. We start simple by building an hour-by-hour simulator, in which we distinguish between 3 types of weather as process states:

- S_1: clear sky
- S_2: cloudy (but dry)
- S_3: rainy

We formulate the probabilities of the future state $\boldsymbol{S}^{n+1} = \{S_1^{n+1}, S_2^{n+1}, S_3^{n+1}\}$ at time $n+1$ (not an exponent) using the following equation:

$p(\boldsymbol{S}^{n+1}) = \mathbf{T} \cdot p(\boldsymbol{S}^n)$

Using historic weather data for the same month as we are interested in, we can conclude the following:

- whenever the sky is clear, it is cloudy the next hour 5% of the time, and it is raining the next hour 3% of the time.
- when the sky is cloudy, 20% of the time we have clear sky the next hour and in 10% we have rain the next hour.
- when it is rainy, 20% of the time we have clear sky the next hour, and 30% of the time it is cloudy.

While having breakfast between 7:00 and 8:00 you observe a clear sky.

3p **9a** Based on the observations above, build the Markov transition matrix $\mathbf{T}$ for this given weather data.

3p **9b** Evaluate the probability that it rains between 9:00 and 10:00.

5p **9c** Explain how Monte Carlo simulation can be used to predict the probability that there will be rain at some point in the afternoon (between 12:00 and 18:00), given your observations at breakfast. In your answer, identify the model, the model input and the model output, as well as the place of the Markov chain in this simulation.

5p **9d** Explain how you can evaluate the average number of hours of rain per day using your model.

After a successful model implementation, we want to go a step further and model rain intensity (in mm/hour) as a stochastic process.

2p **9e** Which of the following stochastic processes can be used for this model of rain intensity? (multiple answers possible)

- ☐ Bernoulli processes
- ☐ Poisson processes
- ☐ Markov processes
- ☐ Gaussian processes

This page is left blank intentionally